Comprehensive and Sensitive Microbial Detection Using A Broad Spectrum Detection Microarray

Crystal Jaing

jaing2@llnl.gov

Planetary Protection Knowledge Gaps
NASA Ames Research Center
March 25, 2015

LLNL is a multidisciplinary national security laboratory

Experimental Test Site
(11 miles² near Tracy, CA)

- Established in 1952
- Approximately 6,500 employees
- 7.1 million gross square feet, 684 facilities
- Annual federal budget: ~ $1.5B

International Space Station—Microbial Observatory of Pathogenic Viruses, Bacteria, and Fungi (ISS-MOP) project

- New Space Biology project

- Lawrence Livermore National Lab

- JPL (Venkat)

- NASA Ames (David J. Smith)

- JSC (Satish Mehta)

ISS-MOP database

# Lawrence Livermore Microbial Detection Array (LLMDA), "the everything test", detects >10,000 microbes

Domain	# families	# species	# sequences*
Archaea	29	293	517
Bacteria	265	5367	24945
Fungi	101	265	395
Protozoa	32	117	191
Virus	94	4219	86931
Total	521	**10261**	112979

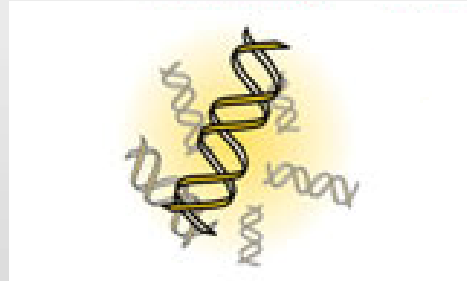*Sequences: complete finished or draft genome, segment, chromosome, plasmid, June of 2013
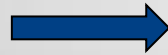
Genomic technologies for pathogen detection
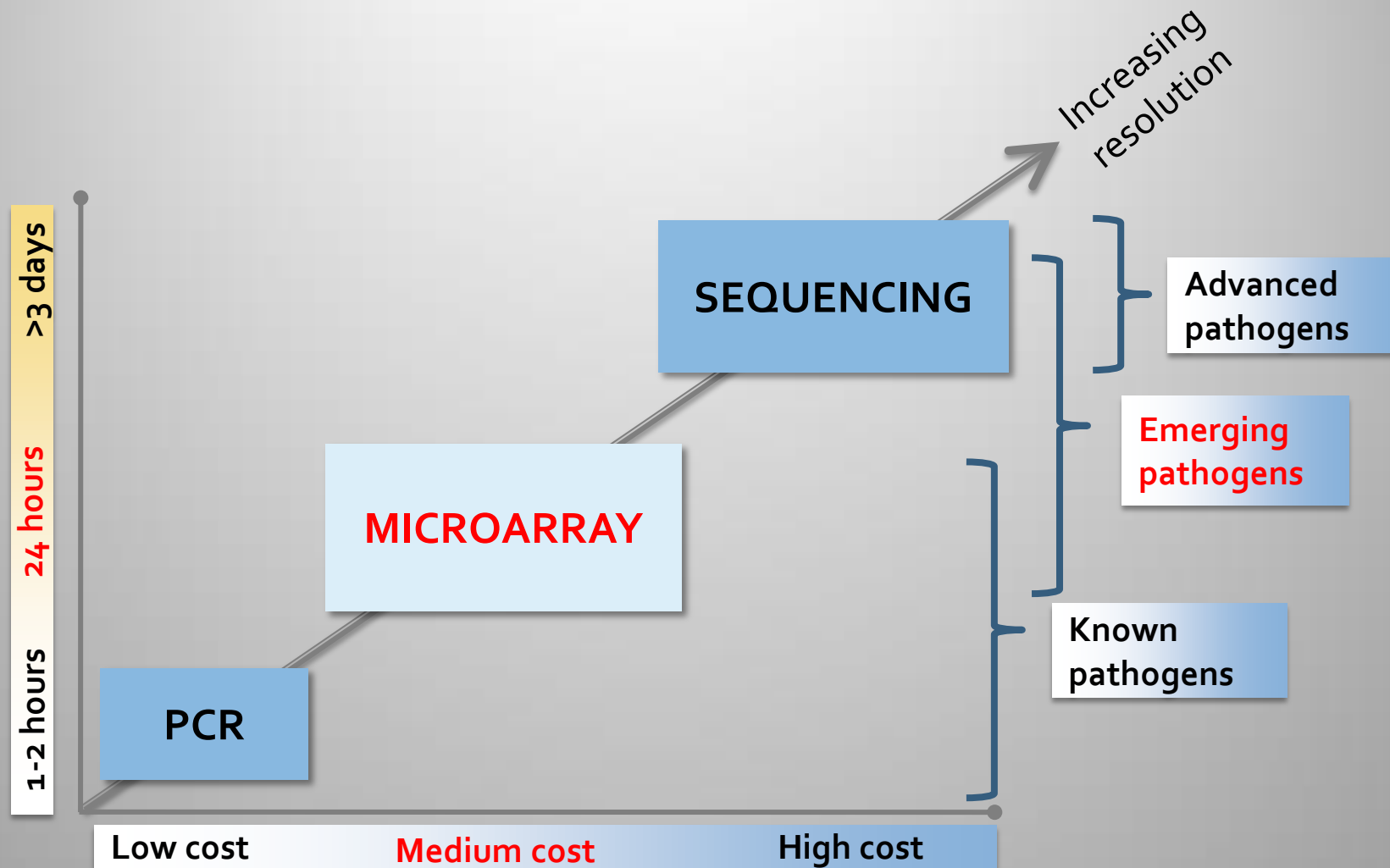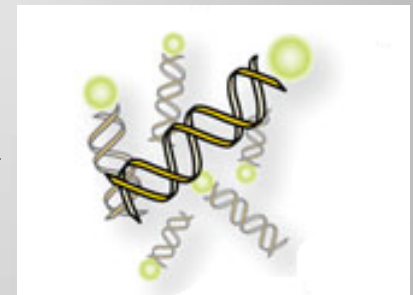
Microarrays can analyze complex clinical and environmental samples
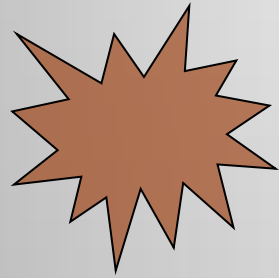
Unknown Sample

Isolate DNA/RNA
Amplification if needed
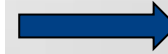
Label with
fluorescent dye

Data analysis

Detect signal on
fluorescent scanner
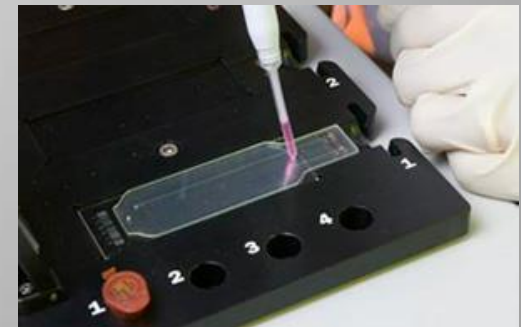
Hybridize sample on array
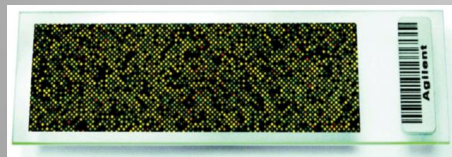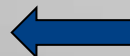
Automated microarray analysis with web data entry, easy to upload

• **Results in ~ 20 minutes**
• **Linux based, requires 64 GB memory**

LLMDA has a wide array of applications

- Environmental monitoring
- Public health
- Vaccine safety
- Food safety
- Biodefense
- Animal health
- Planetary Protection
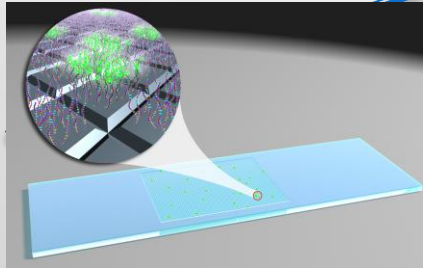
LLMDA has many collaborations
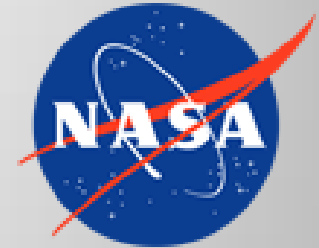
LLMDA in vaccine safety: detected an adventitious pig virus in a vaccine

- **Eight live attenuated vaccines tested, in collaboration with Eric Delwart**

- **Identified a contaminating porcine circovirus from Rotarix vaccine**

- **Being evaluated as risk assessment platform by pharmaceuticals**
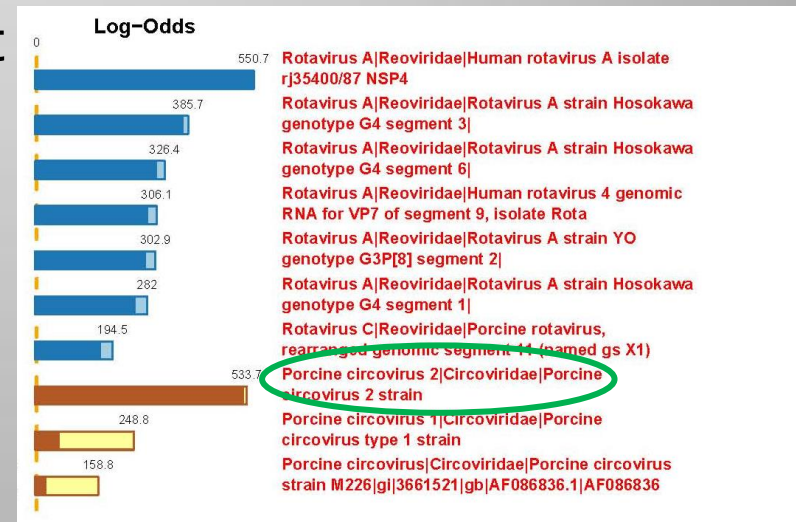
Victoria JG, et al., *J. Virol*. 84: 6033-40, 2010

Log-Odds

550.7		Rotavirus A|Reoviridae|Human rotavirus A isolate rj35400/87 NSP4
385.7		Rotavirus A|Reoviridae|Rotavirus A strain Hosokawa genotype G4 segment 3|
326.4		Rotavirus A|Reoviridae|Rotavirus A strain Hosokawa genotype G4 segment 6|
306.1		Rotavirus A|Reoviridae|Human rotavirus 4 genomic RNA for VP7 of segment 9, isolate Rota
302.9		Rotavirus A|Reoviridae|Rotavirus A strain YO genotype G3P[8] segment 2|
282		Rotavirus A|Reoviridae|Rotavirus A strain Hosokawa genotype G4 segment 1|
194.5		Rotavirus C|Reoviridae|Porcine rotavirus, rearranged genomic segment 11 (named gs X1)
533.7		Porcine circovirus 2|Circoviridae|Porcine circovirus 2 strain
248.8		Porcine circovirus 1|Circoviridae|Porcine circovirus type 1 strain
158.8		Porcine circovirus|Circoviridae|Porcine circovirus strain M226|gi|3661521|gb|AF086836.1|AF086836

LLMDA in infectious disease: identification of wound microbes

GF:733174 (Acinetobacter sp. 6014059 Acinetobacter sp. 6014059, whole genome shotgun sequencing

Acinetobacter sp. SUN plasmid pRAY

Acinetobacter baumannii ACICU plasmid pACICU2

Acinetobacter baumannii plasmid pABIR
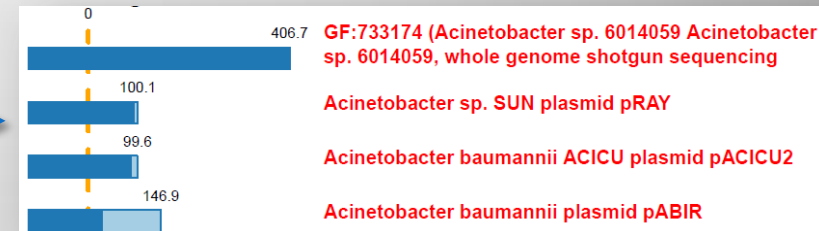
- >100 combat wounded samples (serum, tissue, effluent) analyzed by LLMDA
- More sensitive than culture
- Faster than DNA sequencing

LLMDA identified a range of wound pathogens

Microorganism	Number of positive samples
Acinetobacter sp.	30
Acinetobacter baumannii	28
Escherichia coli	7
Klebsiella pneumoniae	6
Borrelia afzelii	5
Enterobacter cloacae	5
HPV 57	5
HPV 71	5
HHV 6A	4
Salmonella enterica	4
Bacteroides fragilis	3
Human parvovirus	3
Mycobacterium abscessus	3
Pseudomonas sp.	3
Bacillus cereus	2
Pseudomonas aeruginosa	2
Ralstonia solanacearum	2
Staphylococcus lugdunensis	2
Streptomyces sp.	2
Uncultured bacterium plasmid pB10	2

- 50% yielded > 1 detected organism upon analysis

- 37% culture-negative samples are positive via array

- Observed different pathogens associated with failed vs healed wounds

Be et al., 2014, J. Clinical Micro

Secrets in the Bones: The Hunt for the Black Death Killer

Hendrik Poinar, McMaster University

Alison Devault

Identification of *Y. pestis* and *V. cholerae* from ancient DNA samples

- LLMDA detected *Yersinia pestis* in a tooth sample from a medieval Black Death plague corpse dating from 1348

- LLMDA detected *Vibrio cholerae* in a preserved museum specimen from a 1849 cholera victim
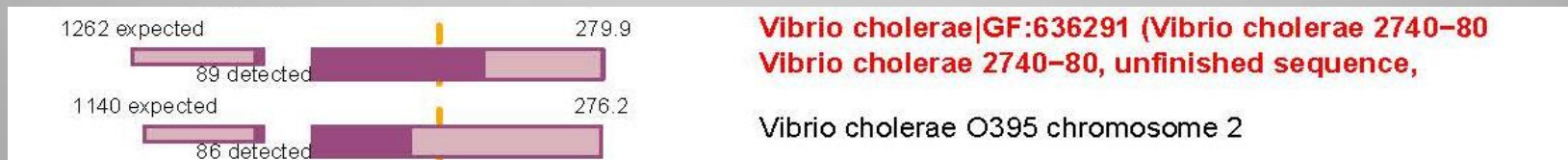
- 1,272 (0.001% of BLAST reads) were *Y. pestis*
- 10,379 (0.009% of BLAST reads) were *V. cholerae*

Devault et al., Scientific Reports, 2014

LLMDA used on ISS dust filter samples

Saliva sample analyzed on LLMDA to assess microbiome and pathogens

Log CL Ratio

0

845.8 — Streptococcus pneumoniae PNI0010

711.7 — Streptococcus sp GMD1S

770.6 — Streptococcus mitis SK575

707.5 — Prevotella denticola CRIS 18C–A

509 — Prevotella intermedia 17 chrchromosome

501.6 — Prevotella marshii DSM 16973

535.7 — Haemophilus parainfluenzae T3T1

474.2 — Haemophilus parainfluenzae HK2019

365.4 — Haemophilus pittmaniae HK 85

451 — Neisseria flavescens SK114

57.8 — Neisseria gonorrhoeae plasmidpJD4

297.1 — Veillonella dispar ATCC 17748

122.6 — Megasphaera micronuciformis F0359

261.9 — Campylobacter concisus 13826

84.7 — Campylobacter concisus 13826 plasmidpCCON16

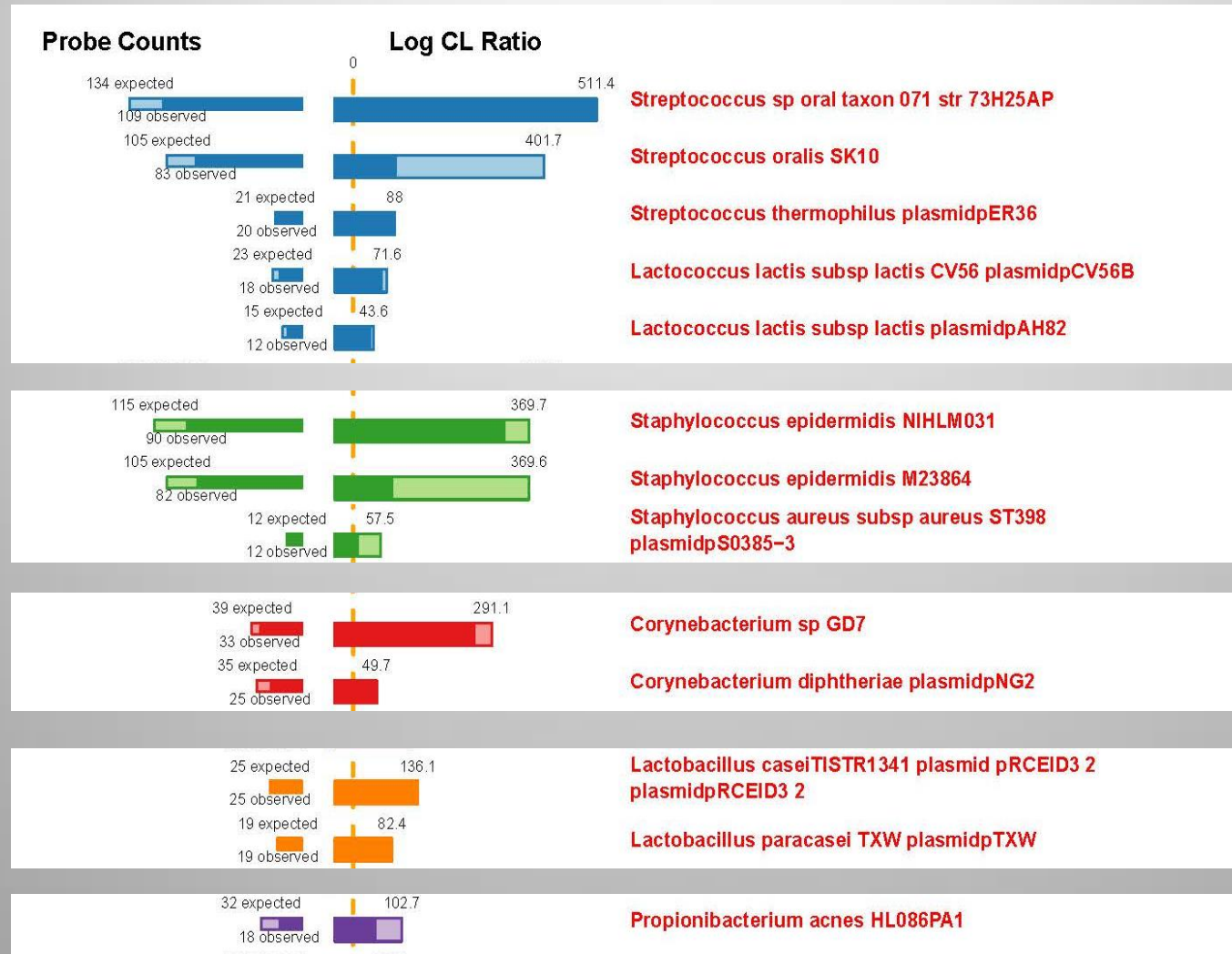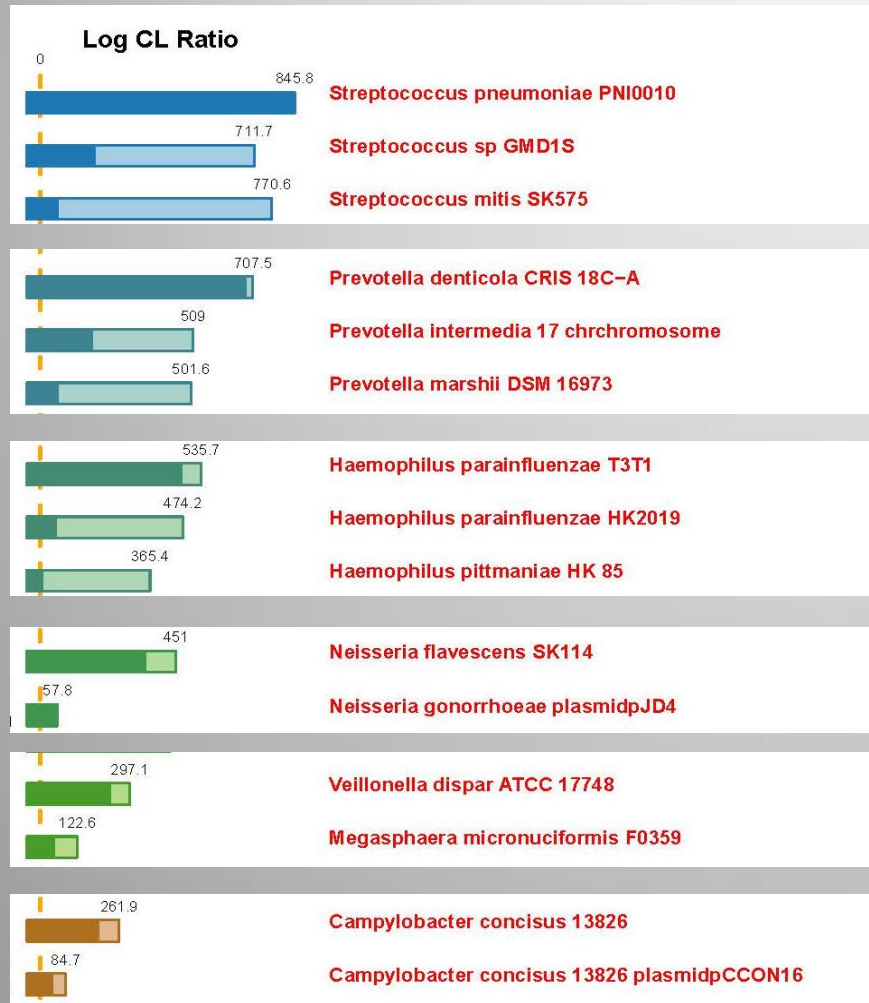271.6 — Rothia mucilaginosa M508

216.5 — Rothia aeria F0474

239.8 — Oribacterium sp ACB7

229.1 — Treponema medium ATCC 700293

201.6 — Fusobacterium nucleatum subsp vincentii ATCC 49256

229.6 — Porphyromonas endodontalis ATCC 35406

215.3 — Bacteroidetes oral taxon 274 str F0058

183.4 — Gemella morbillorum M424

136.6 — Eubacterium saphenum ATCC 49989

107 — Filifactor alocis ATCC 35896

24.9 — Streptococcus phage Cp–1

168.1 — Granulicatella elegans ATCC 700633

11.4 — Human herpesvirus 7

10.7 — Human herpesvirus 7

10.8 — Staphylococcus epidermidis ATCC 12228 plasmidpSE–12228–06

Virulence Chip will be used to characterize virulence, antibiotic resistance of ISS samples
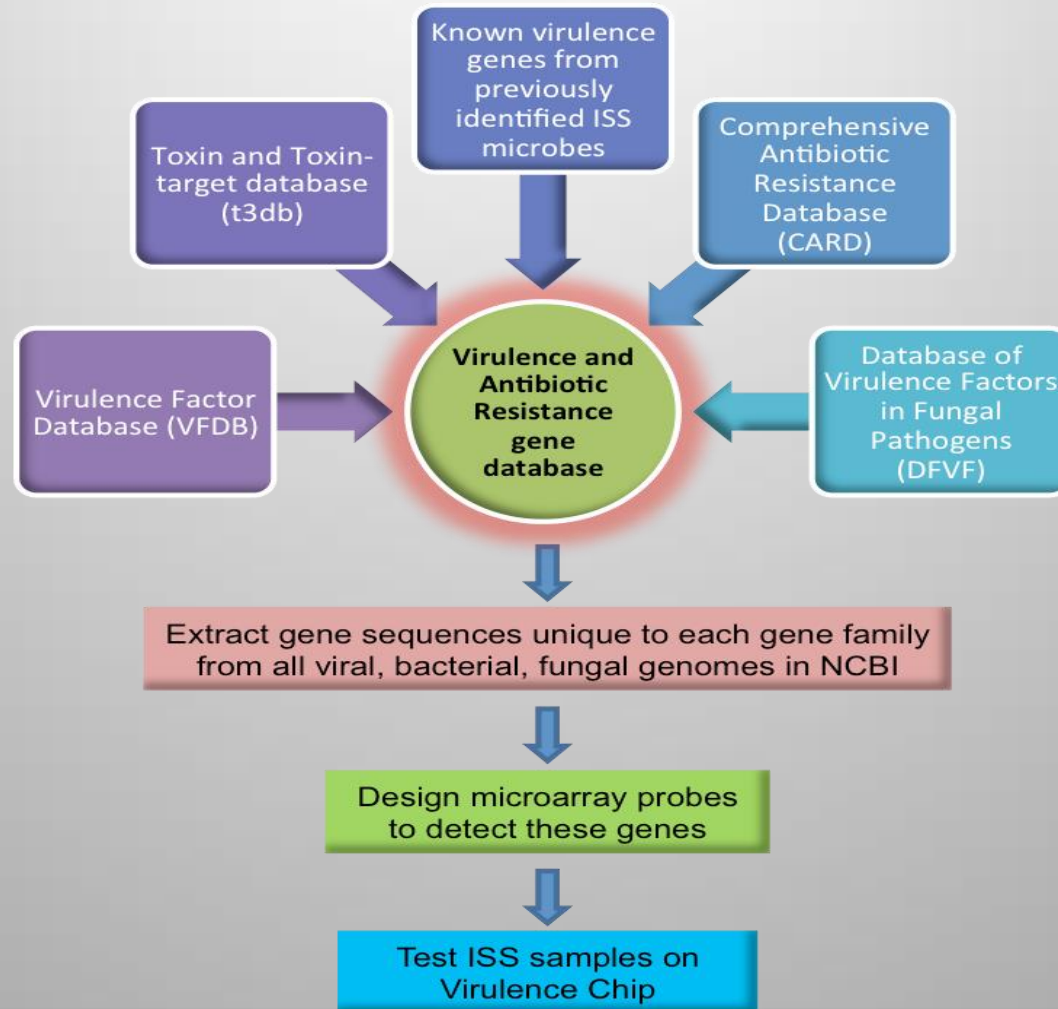
High-Performance Computing (HPC) is part of the Laboratory's DNA

Sequoia ranked #1 on the three lists, set a record by simulating 504 billion events per second.
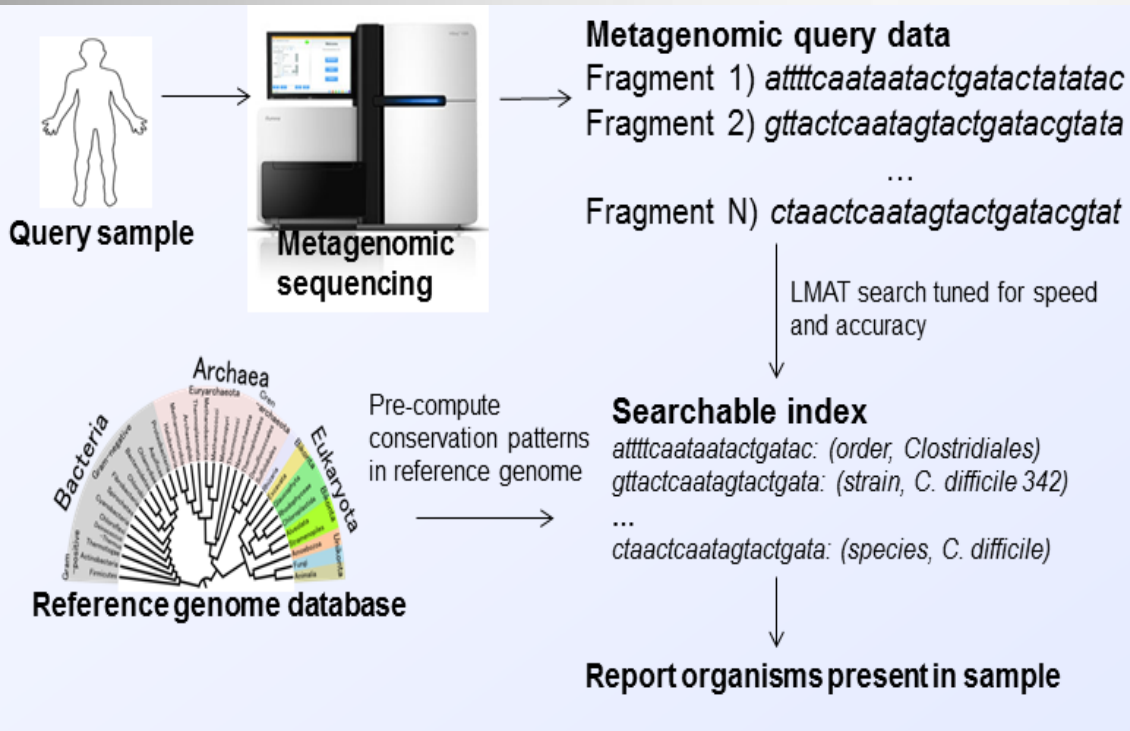
Livermore Metagenomic Analysis Toolkit: shotgun metagenomic analysis conducted on an unprecedented scale

Metagenomic query data
Fragment 1) *attttcaataatactgatactatatac*
Fragment 2) *gttactcaatagtactgatacgtata*
...
Fragment N) *ctaactcaatagtactgatacgtat*

LMAT search tuned for speed and accuracy

Pre-compute conservation patterns in reference genome

Searchable index
attttcaataatactgatac: (order, Clostridiales)
gttactcaatagtactgata: (strain, C. difficile 342)
...
ctaactcaatagtactgata: (species, C. difficile)

Report organisms present in sample

Search all complete and draft genomes:

- viruses, bacteria, archaea
- protozoa, fungi
- human, mitochondrial DNA of larger eukaryotes
- artificial sequence (vectors, etc)
- all sequenced genes

12,632 species

~3 times larger than the closest searchable DB

Application: strain/gene identification and organism discovery

Search many more genes and genomes than otherwise possible using new large memory computing architectures

Future research to develop a fully integrated system from sample prep to pathogen ID